

Patent Application US/07/923,692

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Donson, Jon
Dawson, William O.
Grantham, George L.
Turpen, Thomas H.
Turpen, Ann Myers
Garger, Stephen J.
Grill, Laurence K.

(ii) TITLE OF INVENTION: RECOMBINANT PLANT VIRAL NUCLEIC ACIDS

(iii) NUMBER OF SEQUENCES: 11

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Limbach & Limbach
(B) STREET: 2001 Ferry Building
(C) CITY: San Francisco
(D) STATE: CAL
(F) ZIP: 94111

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Patent in Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 600,244
(B) FILING DATE: 22-OCT-1990

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 641,617
(B) FILING DATE: 16-JAN-1991

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 310,881
(B) FILING DATE: 17-FEB-1989

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 160,766
(B) FILING DATE: 26-FEB-1988

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 160,771
(B) FILING DATE: 26-FEB-1988

Drour
see pp. 5, 6, 7, 14, 15,
and 17.

** In addition, this*
File was not

submitted
in ASCII
text. Please
ensure that
your corrected
disk has an
ASCII file
on it.

Patent Application US/07/923,692

53
54 (vii) PRIOR APPLICATION DATA:
55 (A) APPLICATION NUMBER: US 347,637
56 (B) FILING DATE: 05-MAY-1989
57
58 (vii) PRIOR APPLICATION DATA:
59 (A) APPLICATION NUMBER: US 363,138
60 (B) FILING DATE: 08-JUN-1989
61
62 (vii) PRIOR APPLICATION DATA:
63 (A) APPLICATION NUMBER: US 219,279
64 (B) FILING DATE: 15-JUL-1988
65
66 (viii) ATTORNEY/AGENT INFORMATION:
67 (A) NAME: Halluin, Albert P.
68 (B) REGISTRATION NUMBER: 28,957
69 (C) REFERENCE/DOCKET NUMBER: BIOG-20121 USA
70
71 (ix) TELECOMMUNICATION INFORMATION:
72 (A) TELEPHONE: 415-433-4150
73 (B) TELEFAX: 415-433-8716
74
75
76 (2) INFORMATION FOR SEQ ID NO: 1:
77
78 (i) SEQUENCE CHARACTERISTICS:
79 (A) LENGTH: 4 amino acids
80 (B) TYPE: amino acid
81 (D) TOPOLOGY: linear
82
83 (ii) MOLECULE TYPE: peptide
84
85 (iii) HYPOTHETICAL: NO
86
87 (iv) ANTI-SENSE: NO
88
89 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
90
91 Pro Xaa Gly Pro
92 1
93
94 (2) INFORMATION FOR SEQ ID NO: 2:
95
96 (i) SEQUENCE CHARACTERISTICS:
97 (A) LENGTH: 13 base pairs
98 (B) TYPE: nucleic acid
99 (C) STRANDEDNESS: single
100 (D) TOPOLOGY: linear
101
102 (ii) MOLECULE TYPE: DNA (genomic)
103
104 (iii) HYPOTHETICAL: NO

Patent Application US/07/923,692

105
106 (iv) ANTI-SENSE: NO
107
108 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
109
110 GGGTACCTGG GCC 13
111
112
113
114 (2) INFORMATION FOR SEQ ID NO: 3:
115
116 (i) SEQUENCE CHARACTERISTICS:
117 (A) LENGTH: 886 base pairs
118 (B) TYPE: nucleic acid
119 (C) STRANDEDNESS: single
120 (D) TOPOLOGY: linear
121
122 (ii) MOLECULE TYPE: DNA (genomic)
123
124 (iii) HYPOTHETICAL: NO
125
126 (iv) ANTI-SENSE: NO
127
128 (vi) ORIGINAL SOURCE:
129 (A) ORGANISM: Chinese cucumber
130
131 (vii) IMMEDIATE SOURCE:
132 (B) CLONE: alpha-trichosanthin
133
134 (ix) FEATURE:
135 (A) NAME/KEY: CDS (B) LOCATION: 8. .877
136 (B) LOCATION: 8. .877
137
138 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
139
140 CTCGAGG ATG ATC AGA TTC TTA GTC CTC TCT TTG CTA ATT CTC ACC CTC 49
141
142 Met Ile Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu
143 1 5 10
144
145 TTC CTA ACA ACT CCT GCT GTG GAG GGC GAT GTT AGC TTC CGT TTA TCA 97
146
147 Phe Leu Thr Thr Pro Ala Val Glu Gly Asp Val Ser Phe Arg Leu Ser
148 15 20 25 30
149
150 GGT GCA ACA AGC AGT TCC TAT GGA GTT TTC ATT TCA AAT CTG AGA AAA 145
151
152 Gly Ala Thr Ser Ser Ser Tyr Gly Val Phe Ile Ser Asn Leu Arg Lys
153 35 40 45
154
155 GCT CTT CCA AAT GAA AGG AAA CTG TAC GAT ATC CCT CTG TTA CGT TCC 193
156

Patent Application US/07/923,692

```

157 Ala Leu Pro Asn Glu Arg Lys Leu Tyr Asp Ile Pro Leu Leu Arg Ser
158          50                      55                      60
159
160 TCT CTT CCA GGT TCT CAA CGC TAC GCA TTG ATC CAT CTC ACA AAT TAC      241
161
162 Ser Leu Pro Gly Ser Gln Arg Tyr Ala Leu Ile His Leu Thr Asn Tyr
163          65                      70                      75
164
165 GCC GAT GAA ACC ATT TCA GTG GCC ATA GAC GTA ACG AAC GTC TAT ATT      289
166
167 Ala Asp Glu Thr Ile Ser Val Ala Ile Asp Val Thr Asn Val Tyr Ile
168          80                      85                      90
169
170 ATG GGA TAT CGC GCT GGC GAT ACA TCC TAT TTT TTC AAC GAG GCT TCT      337
171
172 Met Gly Tyr Arg Ala Gly Asp Thr Ser Tyr Phe Phe Asn Glu Ala Ser
173          95                      100                      105                      110
174
175 GCA ACA GAA GCT GCA AAA TAT GTA TTC AAA GAC GCT ATG CGA AAA GTT      385
176
177 Ala Thr Glu Ala Ala Lys Tyr Val Phe Lys Asp Ala Met Arg Lys Val
178          115                      120                      125
179
180 ACG CTT CCA TAT TCT GGC AAT TAC GAA AGG CTT CAA ACT GCT GCG GGC      433
181
182 Thr Leu Pro Tyr Ser Gly Asn Tyr Glu Arg Leu Gln Thr Ala Ala Gly
183          130                      135                      140
184
185 AAA ATA AGG GAA AAT ATT CCG CTT GGA CTC CCA GCT TTG GAC AGT GCC      481
186
187 Lys Ile Arg Glu Asn Ile Pro Leu Gly Leu Pro Ala Leu Asp Ser Ala
188          145                      150                      155
189
190 ATT ACC ACT TTG TTT TAC TAC AAC GCC AAT TCT GCT GCG TCG GCA CTT      529
191
192 Ile Thr Thr Leu Phe Tyr Tyr Asn Ala Asn Ser Ala Ala Ser Ala Leu
193          160                      165                      170
194
195 ATG GTA CTC ATT CAG TCG ACG TCT GAG GCT GCG AGG TAT AAA TTT ATT      577
196
197 Met Val Leu Ile Gln Ser Thr Ser Glu Ala Ala Arg Tyr Lys Phe Ile
198          175                      180                      185                      190
199
200 GAG CAA CAA ATT GGG AAG CGC GTT GAC AAA ACC TTC CTA CCA AGT TTA      625
201
202 Glu Gln Gln Ile Gly Lys Arg Val Asp Lys Thr Phe Leu Pro Ser Leu
203          195                      200                      205
204
205 GCA ATT ATA AGT TTG GAA AAT AGT TGG TCT GCT CTC TCC AAG CAA ATT      673
206
207 Ala Ile Ile Ser Leu Glu Asn Ser Trp Ser Ala Leu Ser Lys Gln Ile
208          210                      215                      220

```

Patent Application US/07/923,692

209
210 CAG ATA GCG AGT ACT AAT AAT GGA CAG TTT GAA ACT CCT GTT GTG CTT 721
211
212 Gln Ile Ala Ser Thr Asn Asn Gly Gln Phe Glu Thr Pro Val Val Leu
213 225 230 235
214
215 ATA AAT GCT CAA AAC CAA CGA GTC ATG ATA ACC AAT GTT GAT GCT GGA 769
216
217 Ile Asn Ala Gln Asn Gln Arg Val Met Ile Thr Asn Val Asp Ala Gly
218 240 245 250
219
220 GTT GTA ACC TCC AAC ATC GCG TTG CTG CTG AAT CGA AAC AAT ATG GCA 817
221
222 Val Val Thr Ser Asn Ile Ala Leu Leu Leu Asn Arg Asn Asn Met Ala
223 255 260 265 270
224
225 GCC ATG GAT GAC GAT GTT CCT ATG ACA CAG AGC TTT GGA TGT GGA AGT 865
226
227 Ala Met Asp Asp Asp Val Pro Met Thr Gln Ser Phe Gly Cys Gly Ser
228 275 280 285
229
230 TAT GCT ATT TAGTAACTCG AG 886
231
232 Tyr Ala Ile
233 290
234
235
236 (2) INFORMATION FOR SEQ ID NO:4:
237
238 (i) SEQUENCE CHARACTERISTICS:
239 (A) LENGTH: 289 amino acids
240 (B) TYPE: amino acid
241 (D) TOPOLOGY: linear
242
243 (ii) MOLECULE TYPE: protein
244
245 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
246
247 *Should be Ile*
248 Met *(Ile)* Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu Phe Leu
249 1 5 10 15
250
251 Thr Thr Pro Ala Val Glu Gly Asp Val Ser Phe Arg Leu Ser Gly Ala
252 20 25 30
253
254 Thr Ser Ser Ser Tyr Gly Val Phe Ile Ser Asn Leu Arg Lys Ala Leu
255 35 40 45
256
257 Pro Asn Glu Arg Lys Leu Tyr Asp Ile Pro Leu Leu Arg Ser Ser Leu
258 50 55 60
259
260 Pro Gly Ser Gln Arg Tyr Ala Leu Ile His Leu Thr Asn Tyr Ala Asp

Patent Application US/07/923,692

```

261      65              70              75              80
262
263 Glu Thr Ile Ser Val Ala Ile Asp Val Thr Asn Val Tyr Ile Met Gly
264              85              90              95
265
266 Tyr Arg Ala Gly Asp Thr Ser Tyr Phe Phe Asn Glu Ala Ser Ala Thr
267              100              105              110
268
269 Glu Ala Ala Lys Tyr Val Phe Lys Asp Ala Met Arg Lys Val Thr Leu
270              115              120              125
271
272 Pro Tyr Ser Gly Asn Tyr Glu Arg Leu Gln Thr Ala Ala Gly Lys Ile
273              130              135              140
274
275 Arg Glu Asn Ile Pro Leu Gly Leu Pro Ala Leu Asp Ser Ala Ile Thr
276 145              150              155              160
277
278 Thr Leu Phe Tyr Tyr Asn Ala Asn Ser Ala Ala Ser Ala Leu Met Val
279              165              170              175
280
281 Leu Ile Gln Ser Thr Ser Glu Ala Ala Arg Tyr Lys Phe Ile Glu Gln
282              180              185              190
283
284 Gln Ile Gly Lys Arg Val Asp Lys Thr Phe Leu Pro Ser Leu Ala Ile
285              195              200              205
286
287 Ile Ser Leu Glu Asn Ser Trp Ser Ala Leu Ser Lys Gln Ile Gln Ile
288              210              215              220
289
290 Ala Ser Thr Asn Asn Gly Gln Phe Glu Thr Pro Val Val Leu Ile Asn
291 225              230              235              240
292
293 Ala Gln Asn Gln Arg Val Met Ile Thr Asn Val Asp Ala Gly Val Val
294              245              250              255
295
296 Thr Ser Asn Ile Ala Leu Leu Leu Asn Arg Asn Asn Met Ala Ala Met
297              260              265              270
298
299 Asp Asp Asp Val Pro Met Thr Gln Ser Phe Gly Cys Gly Ser Tyr Ala
300              275              280              285
301
302 Ile
303
304

```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1452 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

re-check numbering

Patent Application US/07/923,692

```

313 (ii) MOLECULE TYPE: DNA (genomic)
314
315 (iii) HYPOTHETICAL: NO
316
317 (iv) ANTI-SENSE: NO
318
319 (vi) ORIGINAL SOURCE:
320 (A) ORGANISM: Oryza sativa
321
322 (vii) IMMEDIATE SOURCE:
323 (B) CLONE: alpha-amylase
324
325 (ix) FEATURE:
326 (A) NAME/KEY: CDS (B) LOCATION: 12. .1316
327 (B) LOCATION: 12. .1316
328
329 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
330
331 CCTCGAGGTG C ATG CAG GTG CTG AAC ACC ATG GTG AAC A CAC TTC TTG
332
333 Met Gln Val Leu Asn Thr Met Val Asn Lys His Phe Leu
334 1 5 10
335
336 TCC CTT TCG GTC CTC ATC GTC CTC CTT GGC CTC TCC TCC AAC TTG ACA 98
337
338 Ser Leu Ser Val Leu Ile Val Leu Leu Gly Leu Ser Ser Asn Leu Thr
339 15 20 25
340
341 GCC GGG CAA GTC CTG TTT CAG GGA TTC AAC TGG GAG TCG TGG AAG GAG 146
342
343 Ala Gly Gln Val Leu Phe Gln Gly Phe Asn Trp Glu Ser Trp Lys Glu
344 30 35 40 45
345
346 AAT GGC GGG TGG TAC AAC TTC CTG ATG GGC AAG GTG GAC GAC ATC GCC 194
347
348 Asn Gly Gly Trp Tyr Asn Phe Leu Met Gly Lys Val Asp Asp Ile Ala
349 50 55 60
350
351 GCA GCC GGC ATC ACC CAC GTC TGG CTC CCT CCG CCG TCT CAC TCT GTC 242
352
353 Ala Ala Gly Ile Thr His Val Trp Leu Pro Pro Pro Ser His Ser Val
354 65 70 75
355
356 GGC GAG CAA GGC TAC ATG CCT GGG CGG CTG TAC GAT CTG GAC GCG TCT 290
357
358 Gly Glu Gln Gly Tyr Met Pro Gly Arg Leu Tyr Asp Leu Asp Ala Ser
359 80 85 90
360
361 AAG TAC GGC AAC GAG GCG CAG CTC AAG TCG CTG ATC GAG GCG TTC CAT 338
362
363 Lys Tyr Gly Asn Glu Ala Gln Leu Lys Ser Leu Ile Glu Ala Phe His
364 95 100 105

```

← 50 only
48 are
on this
line.
Re-calculate
numbering

Patent Application US/07/923,692

```

365
366 GGC AAG GGC GTC CAG GTG ATC GCC GAC ATC GTC ATC AAC CAC CGC ACG      386
367
368 Gly Lys Gly Val Gln Val Ile Ala Asp Ile Val Ile Asn His Arg Thr
369 110                      115                      120                      125
370
371 GCG GAG CAC AAG GAC GGC CGC GGC ATC TAC TGC CTC TTC GAG GGC GGG      434
372
373 Ala Glu His Lys Asp Gly Arg Gly Ile Tyr Cys Leu Phe Glu Gly Gly
374                      130                      135                      140
375
376 ACG CCC GAC TCC CGC CTC GAC TGG GGC CCG CAC ATG ATC TGC CGC GAC      482
377
378 Thr Pro Asp Ser Arg Leu Asp Trp Gly Pro His Met Ile Cys Arg Asp
379                      145                      150                      155
380
381 GAC CCC TAC GGC CAT GGC ACC GGC AAC CCG GAC ACC GGC GCC GAC TTC      530
382
383 Asp Pro Tyr Gly Asp Gly Thr Gly Asn Pro Asp Thr Gly Ala Asp Phe
384                      160                      165                      170
385
386 GCC GCC GCG CCG GAC ATC GAC CAC CTC AAC AAG CGC GTC CAG CGG GAG      578
387
388 Ala Ala Ala Pro Asp Ile Asp His Leu Asn Lys Arg Val Gln Arg Glu
389                      175                      180                      185
390
391 CTC ATT GGC TGG CTC GAC TGG CTC AAG ATG GAC ATC GGC TTC GAC GCG      626
392
393 Leu Ile Gly Trp Leu Asp Trp Leu Lys Met Asp Ile Gly Phe Asp Ala
394 190                      195                      200                      205
395
396 TGG CGC CTC GAC TTC GCC AAG GGC TAC TCC GCC GAC ATG GCA AAC ATC      674
397
398 Trp Arg Leu Asp Phe Ala Lys Gly Tyr Ser Ala Asp Met Ala Lys Ile
399                      210                      215                      220
400
401 TAC ATC GAC GCC ACC GAG CCG AGC TTC GCC GTG CCC GAG ATA TCG ACG      722
402
403 Tyr Ile Asp Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Thr
404                      225                      230                      235
405
406 TCC ATG GCG AAC GGC GGG GAC GGC AAG CCG AAC TAC GAC CAG AAC GCG      770
407
408 Ser Met Ala Asn Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln Asn Ala
409                      240                      245                      250
410
411 CAC CGG CAG GAG CTG GTC AAC TGG GTC GAT CGT GTC GGC GGC GCC AAC      818
412
413 His Arg Gln Glu Leu Val Asn Trp Val Asp Arg Val Gly Gly Ala Asn
414                      255                      260                      265
415
416 ACC AAC GGC ACG GCG TTC GAC TTC ACC ACC AAG GGC ATC CTC AAC GTC      866

```


Patent Application US/07/923,692

```

417
418 Ser Asn Gly Thr Ala Phe Asp Phe Thr Thr Lys Gly Ile Leu Asn Val
419 270                275                280                285
420
421 GCC GTG GAG GGC GAG CTG TGG CGC CTC CGC GGC GAG GAC GGC AAG GCG      914
422
423 Ala Val Glu Gly Glu Leu Trp Arg Leu Arg Gly Glu Asp Gly Lys Ala
424                290                295                300
425
426 CCC GGC ATG ATC GGG TGC TGG CCG GCC AAG GCG ACG ACC TTC GTC GAC      962
427
428 Pro Gly Met Ile Gly Trp Trp Pro Ala Lys Ala Thr Thr Phe Val Asp
429                305                310                315
430
431 AAC CAC GAC ACC GGC TCG ACG CAG CAC CTG TGG CCG TTC CCC TCC GAC      1010
432
433 Asn His Asp Thr Gly Ser Thr Gln His Leu Trp Pro Phe Pro Ser Asp
434                320                325                330
435
436 AAG GTC ATG CAG GGC TAC GCA TAC ATC CTC ACC CAC CCC GGC AAC CCA      1058
437
438 Lys Val Met Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly Asn Pro
439                335                340                345
440
441 TGC ATC TTG TAC GAC CAT TTC TTC GAT TGG GGT CTC AAG GAG GAG ATC      1106
442
443 Cys Ile Phe Tyr Asp His Phe Phe Asp Trp Gly Leu Lys Glu Glu Ile
444 350                355                360                365
445
446 GAG CGC CTG GTG TCA ATC AGA AAC CGG CAG GGG ATC CAC CCG GCG AGC      1154
447
448 Glu Arg Leu Val Ser Ile Arg Asn Arg Gln Gly Ile His Pro Ala Ser
449                370                375                380
450
451 GAG CTG CGC ATC ATG GAA GCT GAC AGC GAT CTC TAC CTC GCG GAG ATC      1202
452
453 Glu Leu Arg Ile Met Glu Ala Asp Ser Asp Leu Tyr Leu Ala Glu Ile
454                385                390                395
455
456 GAT GGC AAG GTG ATC ACA AAG ATT GGA CCA AGA TAC GAC GTC GAA CAC      1250
457
458 Asp Gly Lys Val Ile Thr Lys Ile Gly Pro Arg Tyr Asp Val Glu His
459                400                405                410
460
461 CTC ATC CCC GAA GGC TTC CAG GTC GTC GCG CAC GGT GAT GGC TAC GCA      1298
462
463 Leu Ile Pro Glu Gly Phe Gln Val Val Ala His Gly Asp Gly Tyr Ala
464 415                420                425
465
466 ATC TGG GAG AAA ATC TGAGCGCACG ATGACGAGAC TCTCAGTTTA GCAGATTTAA      1353
467
468 Ile Trp Glu Lys LIe

```

Patent Application US/07/923,692

469 430 435
470
471 CCTGCGATTT TTACCCTGAC CGGTATACGT ATATACGTGC CGGCAACGAG CTGTATCCGA 1413
472
473
474 TCCGAATTAC GGATGCAATT GTCCACGAAG TCCTCGAGG 1452
475
476
477
478 (2) INFORMATION FOR SEQ ID NO: 6:
479
480 (i) SEQUENCE CHARACTERISTICS:
481 (A) LENGTH: 434 amino acids
482 (B) TYPE: amino acid
483 (D) Topology: linear
484
485 (ii) MOLECULE TYPE: protein
486
487 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
488
489 Met Gln Val Leu Asn Thr Met Val Asn Lys His Phe Leu Ser Leu Ser
490 1 5 10 15
491
492 Val Leu Ile Val Leu Leu Gly Leu Ser Ser Asn Leu Thr Ala Gly Gln
493 20 25 30
494
495 Val Leu Phe Gln Gly Phe Asn Trp Glu Ser Trp Lys Glu Asn Gly Gly
496 35 40 45
497
498 Trp Tyr Asn Phe Leu Met Gly Lys Val Asp Asp Ile Ala Ala Ala Gly
499 50 55 60
500
501 Ile Thr His Val Trp Leu Pro Pro Pro Ser His Ser Val Gly Glu Gln
502 65 70 75 80
503
504 Gly Tyr Met Pro Gly Arg Leu Tyr Asp Leu Asp Ala Ser Lys Tyr Gly
505 85 90 95
506
507 Asn Glu Ala Gln Leu Lys Ser Leu Ile Glu Ala Phe His Gly Lys Gly
508 100 105 110
509
510 Val Gln Val Ile Ala Asp Ile Val Ile Asn His Arg Thr Ala Glu His
511 115 120 125
512
513 Lys Asp Gly Arg Gly Ile Tyr Cys Leu Phe Glu Gly Gly Thr Pro Asp
514 130 135 140
515
516 Ser Arg Leu Asp Trp Gly Pro His Met Ile Cys Arg Asp Asp Pro Tyr
517 145 150 155 160
518
519 Gly Asp Gly Thr Gly Asn Pro Asp Thr Gly Ala Asp Phe Ala Ala Ala
520 165 170 175

Patent Application US/07/923,692

521
522 Pro Asp Ile Asp His Leu Asn Lys Arg Val Gln Arg Glu Leu Ile Gly
523 180 185 190
524
525 Trp Leu Asp Trp Leu Lys Met Asp Ile Gly Phe Asp Ala Trp Arg Leu
526 195 200 205
527
528 Asp Phe Ala Lys Gly Tyr Ser Ala Asp Met Ala Lys Ile Tyr Ile Asp
529 210 215 220
530
531 Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Thr Ser Met Ala
532 225 230 235 240
533
534 Asn Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln Asn Ala His Arg Gln
535 245 250 255
536
537 Glu Leu Val Asn Trp Val Asp Arg Val Gly Gly Ala Asn Ser Asn Gly
538 260 265 270
539
540 Thr Ala Phe Asp Phe Thr Thr Lys Gly Ile Leu Asn Val Ala Val Glu
541 275 280 285
542
543 Gly Glu Leu Trp Arg Leu Arg Gly Glu Asp Gly Lys Ala Pro Gly Met
544 290 295 300
545
546 Ile Gly Trp Trp Pro Ala Lys Ala Thr Thr Phe Val Asp Asn His Asp
547 305 310 315 320
548
549 Thr Gly Ser Thr Gln His Leu Trp Pro Phe Pro Ser Asp Lys Val Met
550 325 330 335
551
552 Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly Asn Pro Cys Ile Phe
553 340 345 350
554
555 Tyr Asp His Phe Phe Asp Trp Gly Leu Lys Glu Glu Ile Glu Arg Leu
556 355 360 365
557
558 Val Ser Ile Arg Asn Arg Gln Gly Ile His Pro Ala Ser Glu Leu Arg
559 370 375 380
560
561 Ile Met Glu Ala Asp Ser Asp Leu Tyr Leu Ala Glu Ile Asp Gly Lys
562 385 390 395 400
563
564 Val Ile Thr Lys Ile Gly Pro Arg Tyr Asp Val Glu His Leu Ile Pro
565 405 410 415
566
567 Glu Gly Phe Gln Val Val Ala His Gly Asp Gly Tyr Ala Ile Trp Glu
568 420 425 430
569
570 Lys Ile
571
572

Patent Application US/07/923,692

573 (2) INFORMATION FOR SEQ ID NO:7:

574

575 (i) SEQUENCE CHARACTERISTICS:

576 (A) LENGTH: 709 base pairs

577 (B) TYPE: nucleic acid

578 (G) STRANDEDNESS: single

579 (D) TOPOLOGY: linear

580

581 (ii) MOLECULE TYPE: cDNA to mRNA

582

583 (iii) HYPOTHETICAL: NO

584

585 (iv) ANTI-SENSE: NO

586

587 (vi) ORIGINAL SOURCE:

588 (A) ORGANISM: Homo sapiens

589

590 (vii) IMMEDIATE SOURCE:

591 (B) CLONE: alpha-hemoglobin

592

593 (ix) FEATURE:

594 (A) NAME/KEY: transit_peptide (B) LOCATION: 26..241

595 (B) LOCATION: 26..241

596

597 (ix) FEATURE:

598 (A) NAME/KEY: CDS

599 (B) LOCATION: 245..670

600

601 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

602

603 CTCGAGGGCA TCTGATCTTT CAAGAATGGC ACAAATTAAC AACATGGCAC AAGGGATACA 60

604

605 AACCCTTAAT CCAATTCCA ATTTCCATAA ACCCCAAGTT CCTAAATCTT CAAGTTTCT 120

606

607 TGTTTTTGA TGAAAAAAC TGAAAATTC AGCAAATTCT ATGTTGGTTT TGAAAAAAGA 180

608

609 TTCAATTTTT ATGCAAAAGT TTTGTTCTT TAGGATTTC GCAGGTGGTA GAGTTTCTTG 240

610

611 CATG GTG CTG TCT CCT GCC GAC AAG ACC AAC GTC AAG GCC GCC TGG GGC 289

612

613 Val Leu Ser Pro Ala Asp Lys Thr Asn Val Lys Ala Ala Trp Cys

614 1 5 10 15

615

616 AAG GTT GGC GCG CAC GCT GGC GAG TAT GGT GCG GAG GCC CTG GAG AGG 337

617

618 Lys Val Gly Ala His Ala Gly Glu Tyr Gly Ala Glu Ala Leu Glu Arg

619 20 25 30

620

621 ATG TTC CTG TCC TTC CCC ACC ACC AAG ACC TAC TTC CCG CAC TTC GAC 385

622

623 Met Phe Leu Ser Phe Pro Thr Thr Lys Thr Tyr Phe Pro His Phe Asp

624 35 40 45

Patent Application US/07/923,692

625
626 CTG AGC CAC GGC TCT GCC CAG GTT AAG GGC CAC GGC AAG AAG GTG GCC 433
627
628 Leu Ser His Gly Ser Ala Gln Val Lys Gly His Gly Lys Lys Val Ala
629 50 55 60
630
631 GAC GCG CTG ACC AAC GCC GTG GCG CAC GTG GAC GAC ATG CCC AAC GCG 481
632
633 Asp Ala Leu Thr Asn Ala Val Ala His Val Asp Asp Met Pro Asn Ala
634 65 70 75
635
636 CTG TCC GCC CTG AGC GAC CTG CAC GCG CAC AAG CTT CGG GTG GAC CCG 529
637
638 Leu Ser Ala Leu Ser Asp Leu His Ala His Lys Leu Arg Val Asp Pro
639 80 85 90 95
640
641 GTC AAC TTC AAG CTC CTA AGC CAC TGC CTG CTG GTG ACC CTG GCC GCC 577
642
643 Val Asn Phe Lys Leu Leu Ser His Cys Leu Leu Val Thr Leu Ala Ala
644 100 105 110
645
646 CAC CTC CCC GCC GAG TTC ACC CCT GCG GTG CAC GCC TCC CTG GAC AAG 625
647
648 His Leu Pro Ala Glu Phe Thr Pro Ala Val His Ala Ser Leu Asp Lys
649 115 120 125
650
651 TTC CTG GCT TCT GTG AGC ACC GTG CTG ACC TCC AAA TAC CGT TAAGCTGGAG 677
652
653 Phe Leu Ala Ser Val Ser Thr Val Leu Thr Ser Lys Tyr Arg
654 130 135 140
655
656
657 CCTCGGTAGC CGTTCCTCCT GCCCGGTCGA CC
658
659
660 (2) INFORMATION FOR SEQ ID NO:8:
661
662 (i) SEQUENCE CHARACTERISTICS:
663 (A) LENGTH: 141 amino acids
664 (B) TYPE: amino acid
665 (D) TOPOLOGY: linear
666
667 (ii) MOLECULE TYPE: protein
668
669
670 (ix) SEQUENCE DESCRIPTION: SEQ ID NO:8:
671
672 Val Leu Ser Pro Ala Asp Lys Thr Asn Val Lys Ala Ala Trp Gly Lys
673 1 5 10 15
674
675 Val Gly Ala His Ala Gly Glu Tyr Gly Ala Glu Ala Leu Glu Arg Met
676 20 25 30

↙ a nucleic number
is required at
the end of this
line.

Patent Application US/07/923,692

677
678 Phe Leu Ser Phe Pro Thr Thr Lys Thr Tyr Phe Pro His Phe Asp Leu
679 35 40 45
680
681 Ser His Gly Ser Ala Gln Val Lys Gly His Gly Lys Lys Val Ala Asp
682 50 55 60
683
684 Ala Leu Thr Asn Ala Val Ala His Val Asp Asp Met Pro Asn Ala Leu
685 65 70 75 80
686
687 Ser Ala Leu Ser Asp Leu His Ala His Lys Leu Arg Val Asp Pro Val
688 85 90 95
689
690 Asn Phe Lys Leu Leu Ser His Cys Leu Leu Val Thr Leu Ala Ala His
691 100 105 110
692
693 Leu Pro Ala Glu Phe Thr Pro Ala Val His Ala Ser Leu Asp Lys Phe
694 115 120 125
695
696 Leu Ala Ser Val Ser Thr Val Leu Thr Ser Lys Tyr Arg
697 130 135 140
698
699

700 (2) INFORMATION FOR SEQ ID NO:9:

701
702 (i) SEQUENCE CHARACTERISTICS:
703 (A) LENGTH: 743 base pairs
704 (B) TYPE: nucleic acid
705 (C) STRANDEDNESS: single
706 (D) TOPOLOGY: linear
707
708 (ii) MOLECULE TYPE: cDNA to mRNA
709
710 (iii) HYPOTHETICAL: NO
711
712 (iv) ANTI-SENSE: NO
713
714 (vi) ORIGINAL SOURCE:
715 (A) ORGANISM: Homo sapiens
716
717 (vii) IMMEDIATE SOURCE:
718 (B) CLONE: beta-hemoglobin
719
720 (ix) FEATURE:
721 (A) NAME/KEY: transit_peptide (B) LOCATION: 26..241
722 (B) LOCATION: 26..241
723
724 (ix) FEATURE:
725 (A) NAME/KEY: CDS
726 (B) LOCATION: 245..685
727
728 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

See next page

Patent Application US/07/923,692

729
 730 CTCGAGGGGA TCTGATCTTT CAAGAATGGC ACAAATTAAC AACATGGCAC AAGGGATACA 60
 731
 732 AACCCTTAAT CCAATTCCA ATTTCCATAA ACCCCAAGTT CCTAAATCTT CAAGTTTTCT 120
 733
 734 TGTTTTTGGG TCTAAAAAAC TGAAAAATTC AGCAAATTCT ATGTTGGTTT TGAAAAAAGA 180
 735
 736 TTCAATTTTT ATGCAAAAAGT TTTGTTCTTT TAGGATTTCA GCAGGTGGTA GAGTTTCTTG 240
 737
 738 GATG GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG TGG 289
 739
 740 Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp
 741 1 5 10 15
 742
 743 GGC AAG GTG AAC GTG GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG 337
 744
 745 Gly Lys Val Asn Val Asp Glu Val Gly Gly Glu Ala Leu Gly Arg Leu
 746 20 25 30
 747
 748 CTG GTG GTC TAC CCT TGG ACC CAG AGG TTC TTT GAG TCC TTT GGG GAT 385
 749
 750 Leu Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp
 751 35 40 45
 752
 753 CTG TCC ACT CCT GAT GCT GTT ATG GGC AAC CCT AAG GTG AAG GCT CAT 433
 754
 755 Leu Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His
 756 50 55 60
 757
 758 GGC AAG AAA GTG CTG GGT GCC TTT AGT GAT GGC CTG GCT CAC CTG GAC 481
 759
 760 Gly Lys Lys Val Leu Gly Ala Phe Ser Asp Gly Leu Ala His Leu Asp
 761 65 70 75
 762 4 ← an extra?
 763 AAC CTC AAG GGC ACC TTT GCC ACCA CTG AGT GAG CTG CAC TGT GAC AAG 529 S 30
 764
 765 Asn Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys
 766 80 85 90 95
 767
 768 CTG CAC GTG GAT CCT GAG AGC TTC AGG CTC CTA GGC AAC GTG CTG GTC 577
 769
 770 Leu His Val Asp Pro Glu Ser Phe Arg Leu Leu Gly Asn Val Leu Val
 771 100 105 110
 772
 773 TGT GTG CTG GCG CAT CAC TTT GGC AAA GAA TTC ACC CCA CCA GTG CAG 625
 774
 775 Cys Val Leu Ala His His Phe Gly Lys Glu Phe Thr Pro Pro Val Gln
 776 115 120 125
 777
 778 GCT GCC TAT CAG AAA GTG GTG GCT GGT GTG GCT AAT GCC CTG GCC CAC 673
 779
 780 Ala Ala Tyr Gln Lys Val Val Ala Gly Val Ala Asn Ala Leu Ala His

Patent Application US/07/923,692

781 130 135 140
782
783 AAG TAT CAC TAAGCTCGCT TTCTTGCTGT CCAATTTCTA TTAAAGGTTT 722
784
785 Lys Tyr His
786 145
787
788 CTTTGTGGGG TCGAGGTCGA C 743
789
790
791
792 (2) INFORMATION FOR SEQ ID NO: 10:
793
794 (i) SEQUENCE CHARACTERISTICS:
795 (A) LENGTH: 146 amino acids
796 (B) TYPE: amino acid
797 (D) TOPOLOGY: linear
798 (ii) MOLECULE TYPE: protein
799
800 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
801
802 Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp Gly
803 1 5 10 15
804
805 Lys Val Asn Val Asp Glu Val Gly Gly Glu Ala Leu Gly Arg Leu Leu
806 20 25 30
807
808 Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp Leu
809 35 40 45
810
811 Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His Gly
812 50 55 60
813
814 Lys Lys Val Leu Gly Ala Phe Ser Asp Gly Leu Ala His Leu Asp Asn
815 65 70 75 80
816
817 Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys Leu
818 85 90 95
819
820 His Val Asp Pro Glu Ser Phe Arg Leu Leu Gly Asn Val Leu Val Cys
821 100 105 110
822
823 Val Leu Ala His His Phe Gly Lys Glu Phe Thr Pro Pro Val Gln Ala
824 115 120 125
825
826 Ala Tyr Gln Lys Val Val Ala Gly Val Ala Asn Ala Leu Ala His Lys
827 130 135 140
828
829 Tyr His
830 145
831
832

Patent Application US/07/923,692

833 (2) INFORMATION FOR SEQ ID NO:11:

834

835 (i) SEQUENCE CHARACTERISTICS:

836 (A) LENGTH: 17 amino acids

837 (B) TYPE: amino acid

838 (D) TOPOLOGY: linear

839

840 (ii) MOLECULE TYPE: peptide

841

842 (v) FRAGMENT TYPE: N-terminal

843

844 (vi) ORIGINAL SOURCE:

845 (A) ORGANISM: alkalophilic Bacillus sp.

846 (B) STRAIN: 38-2

847

848 (vii) IMMEDIATE SOURCE:

849 (B) CLONE: beta-cyclodextrin

850

851 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11.

852

853 Ala Pro Asp Thr Ser Val Ser Asn Lys Gln Asn Phe Ser Thr Asp Val

854 1 5 10 15

855

856 Ile

||
this is an I1.
↑
letter

PAGE: 1

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/07/923,692

DATE: 03/29/93
TIME: 15:41:46
S543

LINE ERROR

ORIGINAL TEXT

248 Wrong Amino Acid Designator
245 Entered and Calc. Seq. Length differ
329 Entered and Calc. Seq. Length differ
601 Entered and Calc. Seq. Length differ
728 Entered and Calc. Seq. Length differ
851 Wrong Sequence Number

Met lIe Arg Phe Leu Val Leu Ser Leu Leu
(xi) SEQUENCE DESCRIPTION: SEQ ID NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO

PAGE: 1

SEQUENCE MISSING ITEM REPORT
PATENT APPLICATION US/07/923,692

DATE: 03/29/93
TIME: 15:41:46
S543

MANDATORY IDENTIFIER THAT WAS NOT FOUND

COUNTRY

PRIOR APPLICATION DATA

More Identifiers Found Than Max Allowed

PAGE: 1

SEQUENCE CORRECTION REPORT
PATENT APPLICATION US/07/923,692

DATE: 03/29/93
TIME: 15:41:46
S543

LINE ORIGINAL TEXT

CORRECTED TEXT

851 (xi) SEQUENCE DESCRIPTION: SEQ ID NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: IL

BACKUP/RESTORE TRANSACTION

Transaction Number:

933

03/29/93

START TIME: 15:22:19

END TIME: 15:42:01

PROCESSING TIME: 00:19:42

Input Set: S543

Translog Code	:	DOSCOPY
Application Serial Number	:	US/07/923,692
National PCT	:	N
Admendment	:	
Application Class	:	435
Application File Dt.	:	19920731